FIELD RESPONSE PREDICTION: FRAMING THE PROBLEM

Material compiled by: Belkis Cabrera-Palmer, Sandia National Laboratories

Date: 2017/06/23

1 TABLE OF CONTENTS

1 EXECUTIVE SUMMARY

Predicting the performance of radiation detection systems at field sites based on measured performance acquired under controlled conditions at test locations, e.g., the Nevada National Security Site (NNSS), remains a standing issue within DNDO's testing methodology. Detector performance is usually evaluated in the performance and operational testing phases, where the measurement configurations are selected to represent radiation source and background configurations of interest to security applications.

We intend to assist DNDO in the process of creating testing procedures that make the testing results meaningful for deployment sites. In particular, we aim at estimating the signal response of a system under test (SUT) at another field site if measured under exactly the same source configuration used during testing. For that, we propose to study how external conditions, like weather, geographical location and site features are expected to affect the system's response at field sites, and to develop a methodology able to provide an estimate of the system's response for given external conditions that could be employed to assess the system's performance when deployed at field sites.

In this manuscript, we have investigated the literature for existing efforts that consider the variability in the detector response at field operations, and have define basic guidelines for a field response prediction methodology. Our intent has been to avoid prescribing to any particular performance metric used by DNDO during testing. Since background estimation is so fundamental to the problem of source detection, we have focused our analysis on considering how to generate detector response data that is representative of the deployment sites. In studying this problem, we have so far reached the following interrelated and preliminary conclusions:

- Estimating the site background will always involve acquiring data representative of the field conditions where the SUT will be deployed.
- Predicting an average background is not enough to demonstrate expected SUT performance if the associated background variance at a given field site is large.
- The correct approach should combine the estimation of the background distribution based on field data collection with a detection algorithm robust against variations within a given range.
- A background-robust algorithm would need to be trained, at a minimum, on data representative of the expected variations at the deployment site.
- A background-robust algorithm used in performance testing would have to be implemented for the SUT's regular field operation in order to realize the predicted field performance.
- An obvious extension of the above ideas is to train the detection algorithm with data collected at the field site where the SUT is deployed. In this regard, the focus of the testing phase could shift from predicting how the SUT will perform at particular field sites to validating the ability of the SUT detection algorithm to predict the correct background expectation given sufficient local data.

2 INTRODUCTION

Predicting the performance of radiation detection systems at field sites based on measured performance acquired under controlled conditions at test locations, e.g., the Nevada National Security Site (NNSS), remains an unsolved and standing issue within DNDO's testing methodology. Detector performance can be defined in terms of the system's ability to detect and/or identify a given source or set of sources, and depends on the signal generated by the detector for the given measurement configuration (i.e., source strength, distance, time, surrounding materials, etc.) and on the quality of the detection algorithm. Detector performance is usually evaluated in the performance and operational testing phases, where the measurement configurations are selected to represent radiation source and background configurations of interest to security applications.

As such, the field performance of a detection system is expected to depend on a variety of factors that modify the source and background contributions to the measured detector signal. The cargo type and the inspection features, representing variations in measurement configuration will likely affect the source and could also affect the background spectra. Weather and environmental conditions, the geographical location and the site features all have a strong effect on the background spectrum but a minimal effect on the source term. Detector operational features like energy resolution, efficiency and calibration drifts (e.g., due to temperature changes), modify the signal at the detection and acquisition level and thus, will have similar effects on both the source and background terms.

This work is intended to assist DNDO in the process of creating testing procedures that make the testing results meaningful for deployment sites. Here, we narrow our scope to focus on the signal response of a detection system at a field site when measured under the same configuration used during testing. As such, variations in cargo type and inspection features are not considered in this work. We focus in the most common signals measured with radiation detection systems, which are gamma energy spectrum and gross counts, while other signals, like source-emission image, are not considered so far in our analysis. Also, our definition of signal response describes the detector output (typically in terms of a spectrum or gross counts) to a given source-background conficuration, and is different form the commonly used technical term "detector response function" employed to derived the behavior of the detector for arbitrary sources.

Even though the detector performance is evaluated from the application of an alarm algorithm to its spectral response, here we do not attempt to predict the actual detector performance based on any given algorithm. We also restrict our scope to fixed or handheld systems for gamma-ray detection used, e.g., at international border crossings to detect illicit trafficking of radioactive materials. The screenings at ports-of-entry (POE) typically comprise primary inspections for alarm generation and secondary inspections for alarm adjudication. Primary inspections are usually performed with Radiation Portal Monitor (RPM) systems based upon polyvinyl toluene (PVT) plastic scintillator, while secondary inspections employ spectroscopic systems, which usually are based upon sodium iodide (NaI) or high purity germanium (HPGe) detectors.

A general and important concept in any detection problem is that the outcome of a detection system will normally fluctuate from measurement to measurement, even under "constant" measurement conditions.

Fluctuations can originate from variations in the underlying physical process (e.g., the emission of radiation). They can also arise due to finite resolution limits of the detection system. Moreover, when the measurement conditions vary (e.g., either due to changes in the environment or in the measurement configuration), it is important to understand how those variations will affect the detection system outcome. Thus, in order to evaluate the detection capability of a system, distributions of the detector output without and with a source present in a given measurement configuration are generally built, corresponding respectively to the background-only p_B and the source-plus-background p_{S+B} distributions. Those distributions are intended to quantize the expected variability in the detection of the given source under the given measurement configuration. By defining a threshold on the detector output—which, for example, is frequently given in multiples of the background distribution's standard deviation—detection is declared if the measurement lies above the threshold.

In the case of radiation detection systems, the p_{S+B} and p_B distributions have minimum width values given by the Poisson fluctuations due to the radiation detection's stochastic nature. When setting any detection threshold, the detection system can make false positive and false negative errors. The frequency of false positive and false negative outputs will depend on the set threshold, and usually enter the system's performance description in terms of Receiver Operating Characteristics (ROC) curves [1]. However, the false positive and false negative frequencies strongly depend on the separation between the count rate distributions which reflects the signal strength, and on the widths of the count rate distributions which reflect the background variance, as illustrated in Figure 1 taken from [2]. In homeland security applications, the difficult but interesting scenarios are the detection of weak or shielded sources, where the relative separation of p_{S+B} and p_b are small. Thus, understanding the background mean and variance is paramount to predicting the system performance in that background environment.

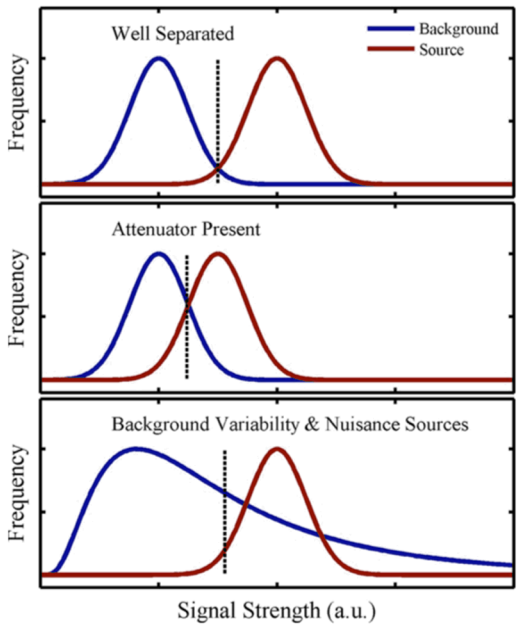

FIGURE 1: SCHEMATIC OF COUNT DISTRIBUTIONS FOR A SOURCE LOCATED IN A STATIC BACKGROUND (TOP), IN THE PRESENCE OF SHIELDING AND/OR BACKGROUND SUPPRESSION (MIDDLE), AND IN THE PRESENCE OF NUISANCE SOURCES (BOTTOM). THE VERTICAL DOTTED LINE REPRESENTS THE DETECTION THRESHOLD. TAKEN FROM [2].

The problem of estimating the expected radiation background, either as gross count or in its spectral form, has received great attention for mobile systems used for radiological source searches since those systems are exposed to a higher background variability as they move through the searched area. Thus, the lessons learned from investigating background variability for mobile systems will directly benefit our study, and we present a summary of some of those methods found in the literature. Even in the case of stationary systems, background variations due to motion can arise from background suppression during vehicle inspection as the cargo moves through the inspection portal and/or from the motion of the handheld system during the inspection of a stopped vehicle. Although those are important effects to study, they are not included in this analysis.

The ultimate goal of this work is twofold: (1) to study how external conditions, like weather, geographical location and site features, and the detector operational conditions are expected to affect the system's signal response at field sites and (2) to propose a methodology that provides an estimate of the system's signal response for given external conditions that could be employed to assess the system's performance when deployed at field sites. The goal of this initial manuscript is more limited: we will investigate the literature for existing efforts that consider the variability in signal response at field operations, and will define basic guidelines for a field response prediction methodology.

Large portions of the material presented in this manuscript have been extracted from the cited references, in some instances intentionally verbatim. This has been done in order to accurately compile the main concepts and ideas as formulated by the researchers of this topic. This document also contains our own analysis, conclusions and ideas on detector field response prediction.

3 GAMMA BACKGROUND VARIABILITY IN THE FIELD

Variation in external factors like environmental conditions, geographical location and site features can modify the radiation background rates non-uniformly across the energy range. Variations in background rate impact the signal-to-noise ratio as a function of energy and thus can independently change the probability of detection for an isotope. In order to understand the origin of gamma background variability, we start with a review of the sources of radioactive background typically found in the field. These are mainly separated into natural, cosmic and anthropogenic sources of radiation. For special nuclear material (SNM) searches, the region of interest (ROI) is usually in the range from 50 keV to 3 MeV. As we will see, most of the background radiation is within that ROI. One exception is when detecting neutrons via emission of excitation gamma-rays with energies beyond 3MeV [3].

3.1 NATURALLY OCCURRING RADIOACTIVE MATERIAL (NORM)

The largest contribution to the natural gamma-ray background originates from the isotopes ^{40}K, ^{238}U, and ^{232}Th through their decay chains. These isotopes are called primordial since they are believed to originate during the early universe and, with half-lives of the order of the age of the earth (5×10^9 years), they are still relatively abundant [4], [5]. Potassium is a major component of the Earth's crust (2.35%) [6], with 0.012% being the radioactive isotope ^{40}K which decays emitting a 1.46 MeV gamma-ray. Uranium is a minor

component of the Earth's crust (~ 3ppm), and occurs naturally as the three unstable isotopes ^{238}U (99.2742%), ^{235}U (0.7204%), and ^{234}U (0.0054%). The most abundant ^{238}U is the parent of a decay series which ends in stable ^{206}Pb, with the most energetic gamma-rays of the series emitted by daughter isotope ^{214}Bi ([7], [8]). Thorium is also a minor component of the Earth's crust (~12 ppm) [6], and occurs naturally only as the unstable radioisotope ^{232}Th which gives rise to a decay series that terminates in the stable isotope ^{208}Pb. Similar to the ^{238}U chain, the ^{232}Th daughter isotope ^{208}Tl emits the most energetic gamma-rays of the series. ^{235}U produces a decay series ending in ^{207}Pb ([7] [8]); however, ^{235}U is less abundant so its contribution to the background is small despite being more radioactive (faster decay) than ^{238}U.

The chemistry for each of these elements dictates their abundance variability in different rock or soil types. In what follows, we present a summary of their chemical properties, directly extracted from [5], [6] and [9], with the only purpose to illustrate how the concentrations of these elements are expected to widely vary across geographical locations according to rock and soil composition. Potassium is found in feldspar[1] minerals and micas, and consequently it is relatively high in felsic rocks (granite, etc.) but low in mafic basalts[2]. Potassium released during rock weathering can be transferred to other K-bearing minerals or adsorbed into clays. The efficient uptake of potassium in clays is reflected in its low concentration in sea water (380 ppm). Thorium occurs in significant quantities in minor minerals such as allanite, monazite, xenotime and zircon. When thorium is weathered out of a mineral it tends to stay in place due to its low solubility. Uranium occurs in many of the same environments, minerals and rocks as thorium when it is in its less-soluble reduced state U^{4+}. Uranium occurs in minor quantities in oxides and silicates (like uraninite and uranothorite respectively) and along grain boundaries. Unlike thorium, uranium also has an oxidized state U^{6+} that is soluble and mobile, but its mobility can be modified by adsorption to hydrous iron oxides, clay minerals and colloids. Thus, the concentration of the primordial radionuclides will depend on the geology of the detector environment [10].

In the U decay series, disequilibrium is common in the natural environment. Disequilibrium occurs when one or more decay products are completely or partially removed or added to the system. Restoring the equilibrium will depend on the half-lives of the radioisotopes involved [4]. Fractioning of the U decay series can be due to both physical and chemical mechanisms. For example, uranium and radium are soluble and thus transportable, and depending on the chemistry of the environment, either uranium or radium can be preferentially leached out of the system relative to the other [4]. For example, ^{226}Ra can be mobilized in high salinity groundwater [6] or can be chemically co-precipitated into oxides or sulfates, adsorbed onto the surface of clays, or adsorbed by plant tissue [4]. These disequilibrium-causing processes add more variability to the natural gamma background, and thus make it more difficult to predict baseline background

---

[1] Feldspars ($KAlSi_3O_8$ – $NaAlSi_3O_8$ – $CaAl_2Si_2O_8$) are a group of rock-forming tectosilicate minerals that make up as much as 60% of the Earth's crust. Taken from Wikipedia.

[2] In geology, felsic refers to igneous rocks that are relatively rich in elements that form feldspar and quartz. It is contrasted with mafic rocks, which are relatively richer in magnesium and iron.

levels based on geological data since the total gamma activity of the uranium daughters might not correlate with the uranium concentration in the ground.

Another important process causing disequilibrium of the U decay series is the escape of the volatile radon gas, specifically the isotope ^{222}Rn. With a half-live of 3.8 days, ^{222}Rn has time to diffuse out of solid material and into the atmosphere [11]. Since the short-lived ^{214}Bi and ^{214}Pb occur below ^{222}Rn in the ^{238}U decay series and are the major gamma-emitters in this series, their contribution to the gamma background from rock and soil will be greatly diminished. However, the background activity from ^{222}Rn and its daughter becomes susceptible to atmospheric changes: as the pressure decreases, ^{222}Rn in the air builds up until it is washed out by the rain [12]. Therefore, the background variability due to ^{222}Rn escape has a time scale given by daily weather conditions.

3.2 Technologically Enhanced Naturally Occurring Radioactive Materials (TENORM)

The low mobility of thorium and its daughter in aqueous environments as well as the short half-lives of its daughters make the Th decay series less susceptible to disequilibrium [3][4]. The radon isotopes occurring in the ^{235}U decay series (^{219}Rn) and ^{232}Th decay series (^{220}Rn) do not tend to cause disequilibrium since their half-lives are very short [4]. However, some ^{220}Rn emanation from ^{232}Th decay series can still be measured in the atmosphere from the activity of its ^{212}Pb daughter, and its corresponding background level follows the same weather dependencies as for ^{222}Rn [11], [12].

Construction materials directly produced from rock and soil contain a wide range of natural radionuclide concentrations reflecting the geological variation of their site of origin ([13] and references therein). Recycled industrial by-products containing Technologically Enhanced Naturally Occurring Radioactive Materials[3] (TENORM) are extensively used in the construction industry and the industrial wastes used in TENORM as substitute for natural products tend to have a relatively higher concentration of Natural Occurring Radioactive Materials (NORM). As a result, the presence of buildings and roads will add a gamma-ray background to the already existing local radiation from the geological conditions. As measured with the mobile system of [14], the background variability due to man-made structures (building, bridges, industrial sites) is significant and occurs in a wide range of spatial scales (from few meters to hundreds of meters). Applying the conclusions from [14] to fixed RPMs, indicate that each separate RPM at the same POE would be subject to a different gamma-ray background dictated by its surroundings structures.

3.3 Cosmogenic sources

[3] Technologically Enhanced Naturally Occurring Radioactive Material (TENORM) is defined as, "Naturally occurring radioactive materials" that have been concentrated or exposed to the accessible environment as a result of human activities such as manufacturing, mineral extraction, or water processing. https://www.epa.gov/radiation/technologically-enhanced-naturally-occurring-radioactive-materials-tenorm

Cosmic rays originate from extraterrestrial sources and are composed primarily of protons (90%), with lesser contributions from alpha particles (9%), electrons (1%) and trace amounts of heavier nuclei and gamma rays [15], [16]. Primary cosmic rays are emitted by extragalactic sources and can interact with interstellar matter generating secondary cosmic rays. Within our solar system, the sun can also generate low energy cosmic rays during solar flares. Upon entering the earth's atmosphere, cosmic rays can interact with atmospheric molecules like oxygen and nitrogen creating air showers of subatomic particles including pions, which can quickly decay in secondary showers generating muons, neutrons and other particles.

A shower can generate on the order of 10^9 particles, which are narrowly confined along the path of the incident cosmic ray. The amount of primary cosmic rays and their secondary particles at the earth's surface is a function of latitude, elevation, and the surrounding environment [17]. These secondary particles can decay, emitting gamma-rays, or interact in the environment to contribute to the gamma-ray background. However, these gamma-ray background sources tend to be relatively weak in the energy region impacted by the terrestrial background described above.

Cosmic-rays are responsible for most naturally-occurring neutrons on the earth surface, including a small contribution from muon-induced spallation and primarily neutron-induced spallation, the so-called "ship effect" [17]. As mentioned, these neutrons would have a relatively weak contribution to the gamma-ray background through interaction with other materials. Cosmic muons, generated in the atmosphere in secondary showers, have a mean lifetime of about 2.2 μs, sufficient to reach the earth's surface, leading to a muon flux at sea level of approximately 170 counts/meters^2/second in the United States [15], [16]. Being highly energetic, these muons do not get fully absorbed in most detectors but create ionization signals proportional to their travel path through the detectors' active material. As such, muons do not present a significant background in the energy region of gamma-rays directly emitted from SNM or other radioisotopes of interest for homeland security. On the other hand, muons can create a considerable high-energy background for larger area detectors like PVTs (see section 4.4) when used for detecting neutron-induced high-energy gamma-rays [3]. For such detectors, diurnal variation of muon rates, cosmic neutrons and cosmic-induced gamma-rays ([3], [18], [19]) should be considered for a complete characterization of the background in the high-energy region [3], [15].

3.4 SKYSHINE

The skyshine consists of gamma-rays emitted from the ground at great distances from a detector and scattered by the air, and contributes a significant amount to the total terrestrial gamma-ray background rate. The work in [20] found that the skyshine contributes about 30% to the terrestrial gamma-ray background, and it is given by a smooth distribution (except for X-rays) concentrated at the low-energy end of the energy spectrum below about 400 keV. As explained by these authors, the skyshine contribution to the terrestrial gamma-ray background might be important when inspecting cargo since a moving container or vehicle will block the different contributions in different ways, adding more background variability. This might be particularly important when attempting to use the low energy gamma-ray emissions for SNM identification.

3.5 ANTHROPOGENIC SOURCES

Some man-made isotopes are incorporated into the radiation background as contamination. The most common example are long-lived fission products found in the background as a result of fallout from nuclear weapons tests and nuclear reactor accidents [2], [11], [21]. The radionuclide ^{137}Cs, with a 30-year half-life, can be found as surface contamination across the United States, Europe, Japan as well as other areas of the world ([22], [23]) and thus, is likely part of the local background in a many POEs. In the event of nuclear accidents like the Chernobyl and Fukushima accidents, shorter half-life isotopes like ^{134}Cs, with a 2-year half-life, can also temporarily appear as part of the background. Industrial radiography sources, such as ^{192}Ir, can also create a local background at a POE if present in a nearby industrial building.

3.6 Nuisance Sources

Nuisance sources are not considered part of the local background (unless they are part of a fixed POE structure or present as local contamination) but represent benign sources found in the stream of commerce. These are usually separated into medical and industrial isotopes. Materials transported as cargo that contain large concentrations of the primordial isotopes, or NORM, also represent nuisance sources for portal monitors. Common examples of NORM include fertilizers, granite, abrasives, ceramics, etc. [2].

4 Detector spectral variation due to external field conditions

The previous section described the variability in the ambient gamma-ray background that any gamma-ray detection system will encounter in the field. In this section, we focus on the response of those detection systems to the changes in radiation background. We also address how the detector operational response can be directly affected by changes in environmental conditions like temperature. The detector response is expressed through the energy spectrum, i.e., the event frequency per energy channel, which depends on specific detector features such as the detection mechanism or material, resolution, and efficiency. We will describe how the spectrum changes according to external conditions causing various levels of measured gamma-ray background and/or alteration in the detector operational parameters. We start with a brief description of the interaction of gamma-rays with the detector active and surrounding materials, and continue with the various detection technologies most commonly used in border security. We categorized the detection technologies accoding to the detection material and without referring to any specific detector brand, since most of the detector features relevant for the applications of this work will depend on the detection material.

4.1 Gamma-ray interactions and their spectrum

The main interaction processes of photons below 3MeV with matter—either the detector media or the surrounding materials—are: photoelectric absorption, Compton scattering and pair production [24]. The photoelectric effect results in all the energy of the photon being absorbed by the bound electron of an atom as kinetic energy plus the emission of an X-ray when the electron vacancy is filled. This is the predominant absorption process at low energy levels, with cross-section increasing strongly with atomic number. In Compton scattering, the photon loses part of its energy to an electron and is scattered at an

angle to its original direction. This is the predominant process for moderate energy photons. At higher energies, pair production becomes the dominant process. Photons with energies above 1.022 MeV (twice the mass of an electron), can be completely absorbed within the electrostatic field of a nucleus with the simultaneous creation of an electron-positron pair. The resultant positron will quickly annihilate with an electron producing two 511 keV gamma rays.

In most radiation detectors, the electrons ejected from the gamma interactions will deposit their kinetic energies within the detector's active material producing an ionization or scintillation signal from which a spectrum is obtained. The ultimate spectral profile for each detector will depend on detector's specifics like the active material atomic number, the conversion of deposited energy into an electronic pulse, the electronic noise, etc. But common to all detectors is that the total energy deposited per incident gamma will depend to first order on the detector size. In the case of photoelectric absorption, the photon's entire energy will be deposited within a detector's volume as long as the detector is large enough to stop the electron and the X-ray, contributing to the full energy peak. When Compton scattering occurs, the scattered photon can either be further absorbed in the detector material (via a photoelectric interaction) so that the total energy deposited contributes to the full energy peak, or it can escape the detector. When the latter occurs, only a fraction of the original photon energy is deposited which results in a continuum of energies below the full energy peak. The intensity of this Compton continuum will depend on the detector size since a second photon interaction will be less likely in smaller detectors. In a similar fashion, the two 511 keV gamma rays emitted after the positron annihilation following pair creation will have greater chance of escaping in smaller detectors. In this case, the energy resulting from a pair production event will contribute to either the single- escape, the double-escape or the full energy peak.

Each background gamma-ray source will contribute to the detector background spectra with its own spectral terms—full energy peak, Compton continuum, single- and double-escape peaks, etc.—, proportionally to its strength. These background gamma rays can also Compton scatter before reaching the detector and add a contribution to the energy continuum that is dependent on their origin within the bulk of the source and on the proximity of the detector to surrounding materials. As a result, any non-background signal that one wishes to detect will be sitting on a background continuum—even if sparsely populated for short measurement times—or possibly overlapping with background peaks depending on the detector's resolution.

This complexity of the gamma spectrum measured by a detection system directly impacts source detection performance. For situations with a low signal-to-background ratio, as is typically the case in homeland security applications, the signal detection sensitivity is dependent on accurate estimates of the background. The most common approach for background estimation in field detection operations is to use previously measured background data. There is also extensive work on modelling the backgrounds most commonly found in field applications ([20], [25], [26], [27]) to be used in detector and algorithm testing as possible substitution to intensive background measurement campaigns. Although such models could be complex, they might be still unable to capture in situ spectral changes (e.g., nearby building materials, temperature changes, etc.) relevant when detecting small signals.

4.2 HPGE DETECTORS

High purity germanium detectors ([21], [25], [28], [29]) are the gold standard for gamma-ray spectroscopy, with resolutions of typically 1.3 keV (0.2%) full width half maximum at 662 keV. This allows precise determination of peak energies, separation of peaks and detection of relatively weak peaks in the presence of a strong background; see an example HPGe spectrum in Figure 2. Large volume HPGe detectors are cylindrical, and the geometries more typically found in field applications are coaxial and closed-end coaxial diode configuration, with volumes up to several hundred cubic centimeters, but the more recent point-contact cylindrical geometries are gradually becoming more common [28], [29]. They are often quoted as percent efficiency at 1332 keV relative to a 3x3 inch NaI(Tl) detector, reaching values greater than 100 %.

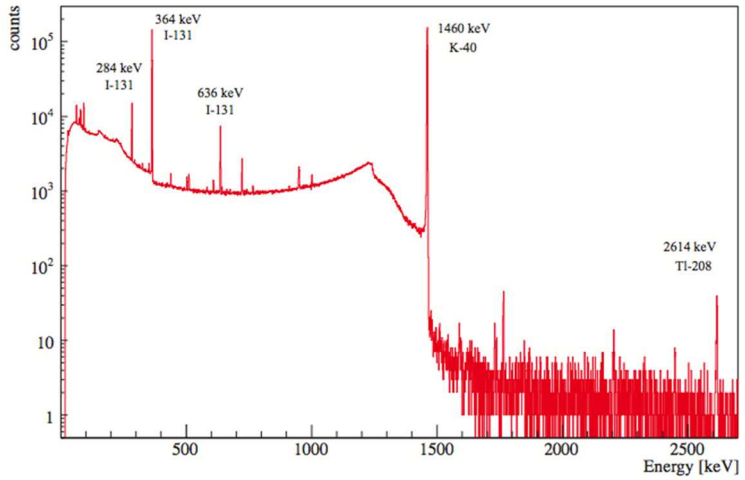

FIGURE 2: HPGE SPECTRUM OF AN ENVIRONMENTAL SAMPLE WHICH IS DOMINATED BY THE COMPTON CONTINUUM FROM THE NATURALLY-OCCURRING ISOTOPE 40K. THE DETECTOR USED WAS A 115% EFFICIENT, N-TYPE COAXIAL DETECTOR IN A LOW-BACKGROUND FACILITY. TAKEN FROM [11].

In the HPGe semiconductor material, the electron-hole pairs created by the gamma-ray-ejected electrons are swept up by an applied voltage to give an electrical signal. As a consequence, HPGe detectors have the disadvantage that they must be operated at low temperatures (less than 100 K) to avoid excessive electronic noise from thermally-induced electron-hole pairs. This is normally achieved by connecting the detector to a dewar containing liquid nitrogen, or by using electro-mechanical coolers, which tend to be more suitable in fixed locations since they only require electrical power. Fixed HPGe systems found in field applications are frequently operated in environmentally controlled cabinets that maintain a stable temperature [30] for the system's electronics and facilitate the Ge crystal cooling. However, systems operated in uncontrolled environmental conditions could be subject to significant temperature variations. The question of whether such ambient temperature variations affect the spectra was studied in [31] for several medium size HPGe detectors. For temperatures in the range from ~5°C to ~30°C [4], no significant

[4] According to [31], HPGe detector manufacturers Ortec and Canberra recommended to only operate their systems within the above range, mainly due to operating limits of the cryo-cooler systems. Though larger temperature ranges can be found in field locations, it might be unreasonable to set up any delicate

changes in resolution and centroid positions were found for gamma-ray peaks spanning the energy range of interest in security applications. Due to the excellent energy resolution of these detectors, small percentage deviations from its nominal resolution will normally be irrelevant.

In HPGe detectors, the excellent energy resolution usually provides the ability to estimate the background contribution within the energy window of the peaks of interest from the neighboring energy windows—unless there is an overlap between the signal of interest and background peaks. Ideally, this eliminates the need to collect a background spectrum in the absence of the inspected object, and thus reduces systematic uncertainties due to possible background changes between those two measurements. However, a higher mean background will also have a higher background distribution variance, increasing the false alarm rate for a given threshold. Thus, even for HPGe detectors understanding of the level of the background continuum is important in order to predict the correct detection sensitivity for a given false alarm rate.

4.3 NaI(Tl) DETECTORS

Thallium activated sodium iodide detectors, NaI(Tl), are commonly used detectors in secondary field inspections; they are conveniently available as hand-held detectors, operate at room temperature, and have high intrinsic detection efficiency. As with all inorganic scintillator, the gamma-ray-induced electrons interact to excite the crystal structure which then decays by emitting photons. These are converted to an electrical signal by the use of photomultiplier tubes (PMT) or photodiodes. NaI(Tl) has large light output but present a limited energy resolution of about 6 - 8% at 662 keV, see Figure 3. This limits its spectroscopic abilities to the identification of only few radionuclides at a time, preferably above a previously known background spectrum. The moderate NaI(Tl) resolution implies that variations in background potentially have a higher impact on isotope detection capability compared to detection with HPGe.

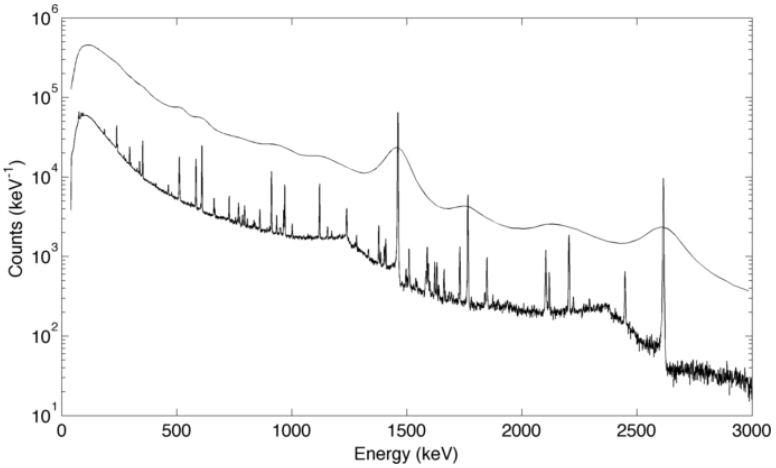

FIGURE 3: COMPARISON OF NAI (TOP CURVE) AND HPGE (BOTTOM CURVE) ENERGY SPECTRA FOR A TYPICAL DAY OF BACKGROUND IN THE BAY AREA (5.7 HOURS TOTAL). TAKEN FROM [11].

---

electronic system in a configuration where it would be directly exposed to very extreme temperature variations.

Variations in ambient temperature modify the NaI(Tl) crystal light output ([20], [24], [32], [33], [34] , [35]), which changes the scintillation signal gain and thus the final spectrum. In [32], two NaI(Tl) detectors were exposed to the range of temperatures from -50°C to +60°C, including the scintillator crystal, the photomultiplier tubes and the preamplifiers. Their results show the maximum signal gain at 0°C, and a total gain variation of ~30% compared to the lowest gain at -50 °C. For more moderate temperature variations within the range 0°C to +20°C, the gain change is much smaller ([32], [34]) but should still be considered when performing spectroscopic measurements. The best resolution was at 20°C (~8% at 662 keV), and changed by about 20% between 20°C and -50°C. These authors recommend that any field application of these detectors should monitor the detector temperature and incorporate adjustments in the data analysis to compensate for the known gain shifts, or alternatively, a constant temperature environment could be maintained for the NaI(Tl) scintillators.

4.4 PVT DETECTORS

Polyvinyl toluene (PVT) scintillator detectors are used in many large-area radiation portal monitor systems. Being an organic scintillator, the gamma-ray-induced electrons generate florescence photons that are converted to an electrical signal by PMTs. Due to its low average atomic number, Compton scattering is the dominantly observed interaction mechanism in PVT detectors between ~20 keV to 20 MeV, the energy range generally of interest for illicit nuclear material, which includes the use of PVT as a neutron detector via neutron-induced gamma-ray emissions [3], [15]. As such, PVT detectors have negligible full-energy peak efficiency and very limited spectroscopic capabilities, see Figure 4, but can be inexpensively built as large area detectors.

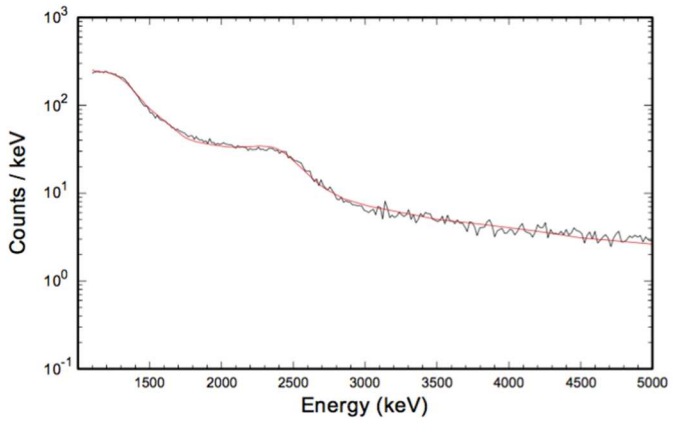

FIGURE 4: A PVT BACKGROUND SPECTRUM TAKEN FROM [18]

Natural background gamma-ray rates observed in such large systems are on the order of thousands of counts per second. The measured diurnal background variations of gamma rays, as observed in deployed RPMs, are illustrated in Figure 5 ([15], [36]) and amount to about 1% from the mean. The total gamma ray oscillations are noted to peak in the late evening hours, and reach a minimum in the late morning hours, and is attributed to diurnal fluctuations in the cosmic-ray intensity ([36]).

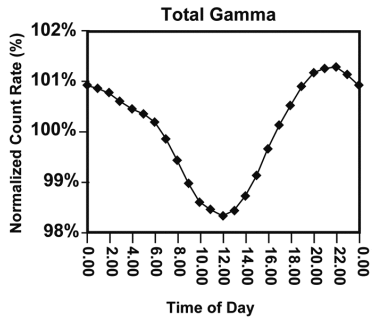

FIGURE 5: AVERAGE DIURNAL RESPONSE OF TOTAL GAMMA-RAY BACKGROUND AS A FUNCTION OF TIME OF DAY, AVERAGED OVER MEASURED RPM PANELS, LOCATIONS, AND DAYS. TAKEN FROM [36].

The study presented in [3] observed that the output from the PVT detectors is relatively insensitive to temperature, but the PMTs exhibit a notable temperature coefficient. Despite the use of coolers regulated to 27 ^{0}C, the authors of [3] observed substantial gain drifts in a 15-hour period, which was attributed to the coolers operation. Though gain adjustment methods are not hard to implement as shown in [3], this constitutes another example highlighting how detector operational variability has a strong impact on detector field performance and should be thoroughly tested and accounted for in any detection system.

5 CURRENT APPROACHES TO ACCOUNT FOR BACKGROUND VARIABILITY

In radiation detection systems, proper characterization of the radiation background is essential in order to maintain high detection sensitivity and low false alarm rates. When background signatures exhibit large variation, alarm thresholds must be set relatively high to void an excess of false alarms, thus reducing sensitivity to source signals [37]. Accounting for background variability is especially important in low signal-to-background ratio scenarios typical in homeland security applications. Such applications are divided into screening for illicit nuclear material at stationary radiation portal monitors and radiological searches with mobile systems. Searches with mobile systems, also referred to as standoff detection, are particularly challenging as the background rate not only varies across different locations and environmental conditions [38], [39], but might not be known or measurable a priori. Hence, there is an extensive and relatively recent body of work devoted to reducing the impact of background variation on alarm rates for mobile detection systems [11], [14], [26], [27]. In the case of stationary RPMs, the exposure to background variability arises from the wide distribution of these systems across the country, as well as from variation in local environmental conditions. In that regard, data and techniques developed for mobile search applications would be not only informative for stationary radiological detection systems, but can also guide the development of algorithms to be employed in system testing to account for the diverse backgrounds in field conditions.

5.1 MAPPING EXTERNAL CONDITIONS TO EXPECTED BACKGROUND SPECTRA FOR MOBILE SYSTEMS

One strategy to mitigate the impact of background variation during search involves extensive characterization of the expected background over time and/or space [37]. Characterizing the expected background requires establishing a baseline of comparison by cataloguing changes in background prior to search and/or tracking those changes during search. As mentioned in [37], one could map the background of geographic regions using global-positioning systems and a decision algorithm can then compare observations during search to this database of prior measurements. But as also stated in [37], environmental and temporal variation may not be fully characterized by such algorithms. Moreover, it is a consensus within the search community that the impact of local conditions on the gamma-ray spectrum can easily overwhelm the radiation background predicted based on geological radioactivity mapping [10].

An effort to characterize the background radiation according to the specific local scene is presented in [14], where various environments were surveyed and categorized based on the distribution of background that they exhibit. These background measurements were performed with the Mobile Imaging and Spectroscopic Threat Identification (MISTI) system, developed by the Naval Research Laboratory, and later renamed as the Radiological Multi-sensor Analysis Platform (RadMAP), which comprises an array of 24 HPGe detectors and an 10x10 (coded-aperture) array of NaI(Tl) detectors. According to [14], if the measured distributions can be correlated to external information about the environment—where such contextual data would be provided by other sensors like cameras, weather stations and GPS—, then the background can be estimated or even modeled when a mobile system moves to a new location, and the alarm threshold adjusted accordingly, providing increased sensitivity (true positive rate) while maintaining specificity (true negative rate). In [14], the measured backgrounds were divided into four categories appropriate for the surveyed San Francisco Bay Area: bridges, rural, downtown and industrial, Figure 6. According to [14], specific categories would be meaningful not only if they have low systematic variability, but also if they are identifiable by some sensors.

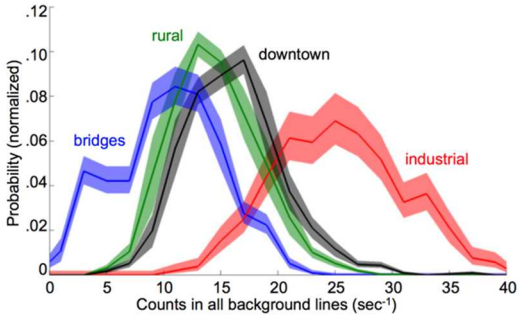

FIGURE 6: BACKGROUND PROBABILITY DISTRIBUTIONS FOR FOUR TYPICAL ENVIRONMENTS: BRIDGES (BLUE), RURAL AREAS (GREEN), DOWNTOWN OAKLAND (BLACK), AND THE JACK LONDON DISTRICT (INDUSTRIAL WAREHOUSES, RED). TAKEN FROM [11].

It is worth noticing that a posterior study presented in [40] using the same data of [14] shows that the background not only adds statistical Poisson noise to the detection, but can also vary systematically as the detector moves from one location to the other producing a much wider systematic uncertainty that should be considered when setting detection thresholds. These considerations are important when using source injection to predict detector performance.

Background modelling is another approach to characterize the expected background spectrum. As mentioned in [14], the information provided by background measurement campaigns can be the input to modelling efforts attempting to predict the background distribution that a given detector system would observe. In [25], a gamma-ray background model for a typical urban environment is created by simulating the response of a typical coaxial HPGe detector with the Monte Carlo-based Geant4 toolkit [41] and matching it with observed data. Though [25] succeeds in satisfactorily approaching the measured data, their multi-stage process reveals the difficulties to account for all background lines and scatter radiation contributing to the final spectra. Another example of background modeling is presented in [20], where the terrestrial background spectrum expected from the three families of isotopes, ^{40}K, ^{232}Th and ^{235}U/^{238}U, measured with a HPGe detector in an open space location at a seaport is modelled using the Monte Caro N-Particle (MCNP) transport code [42]. In contrast with [25], this work manages to match with good agreement the simulation and measured spectra using simple scaling of each family of isotopes; only at energies below 400 keV the difference is about 30%, which the authors discovered is due to lack of accounting for the skyshine effect. Once the skyshine, cosmic and terrestrial gamma backgrounds are considered, the simulation of a NaI(Tl) detector spectra in [20] achieves an excellent match with data collected in an open field site. This work might suggest that modelling the detector spectra in an open site or field, without unknown contributions from nearby structures or buildings, is possible with acceptable accuracy and effort.

Another current effort relevant to search applications combining data collection and modelling is the Multi-agency Urban Search Experiments (MUSE) [26], [27]. Though no peer-reviewed publications on MUSE are available at this time, [26] states that the goal of MUSE is to develop a comprehensive operational radiation transport modeling framework that is validated and quantified against the newly collected benchmark data sets in order to systematically evaluate detector system and algorithm performance. From [27], the synthetic dataset ensemble that MUSE aims to generate and benchmark should cover different background compositions and variability that can be used to test algorithm performance over a variety of parameters.

5.2 Algorithms robust against background variations for mobile systems

A common strategy to account for diverse radiological backgrounds is the use of algorithms that enable background estimation without comparing against a measured background spectrum specific to the measurement scene and moment, or an expected background spectrum based on external conditions. Typically, these detection algorithms train on previously measured background data that is supposed to contain all the variability expected in the measurement scene, so that their outcome (threat/no threat, isotope identification, source location, etc.) is robust against variations in the background. The list presented below is not intended to be comprehensive, but to give the flavor of this approach through some examples. Also, it represents what we have found so far in our literature search, thus important methods unknown to the authors might be unintentionally omitted. In this section, we focus on methods applied to data from mobile systems. In many cases the same methods could be easily adapted to fixed systems like portal monitors or to handheld detectors used in “walkthrough” measurement scenarios.

5.2.1 Region-Of-Interest algorithms

Typical region-of-interest (RoI) algorithms estimate the background under the signal peak using energy windows immediately on the sides of the energy window of interest. Using the NaI detector data collected with RadMAP, [43] presents a modification of the typical RoI method where the background windows are not necessarily chosen next to the photo-peak window when previous background measurements are available. In that case, the estimation of the background is represented to first order by $B = a_0 + a_1B_1+a_2B_2$, where the linear terms a_1 and a_2 represent the contribution from the neighboring windows, and a_0 is the mean excess in the photo-peak window. The counts in the neighbor windows B_1 and B_2 add additional statistical noise, but allow the systematic variance (due to the background variability in mobile searches) to be reduced. The factors a_i are generated using a separate day of background data as a training set to provide the values of B, B_1 and B_2 and fitting them to the above equation using least squares regression. In [43], this background estimation based on training tends to improve performance; however, this effect depends on how well the training set is representative of the full variability encountered in the field.

A similar background estimation technique is described in [44] and [45] under the name Energy Windowing Regression. In their case, the windows selection is based on match filtering to a source template in order to select a range of energies (called the source window W) likely containing the source signal. By linearly regressing the background gamma counts of training data outside the source window (call those energy bins $\overline{W}$), the number of background gamma counts inside the source window W can be predicted. In [44], Least Squares estimator and Ridge Regression estimator are also presented.

5.2.2 Poisson Clutter Split algorithm

According to [46], traditional RoI detection algorithms are unable to effectively account for natural radiological backgrounds and their variation, which has led to the development of more advanced algorithms that are capable of accurate estimation of the background using information from the entire spectrum [2], [47], [48]. In [46], the detection sensitivity from their Poisson Clutter Split (PCS) algorithm [49] is compared to the results from the application of the modified RoI algorithm of [43]. Using source injection on RadMAP background data, [46] shows that PCS improves the probability of detection (for fixed false alarm rate) compared to the RoI algorithm. What makes the PCS algorithm interesting for our proposal is the claim that it can mitigate the two sources of randomness in the radiological spectra: the background clutter, i.e. changes in the energy-dependent count rate due to variations in isotopic composition at different locales, weather conditions, etc..., and the random nature of radioactive decay. According to [46], [49], PCS uses a probabilistic representation of radiological backgrounds, modeling of gamma-ray counts based on Poisson statistics, and a Generalized Likelihood Ratio Test (GLRT) to simultaneously perform detection and identification. The PCS background clutter model encompasses the mean rate spectrum and the dominant and non-linear modes of spectral variation $\vec{b} = \vec{b}(\vec{w})$, and the probability distribution of the coefficients $f(\vec{w})$. The probabilistic model of background rate clutter in combination with the Poisson model for count spectra given the rate is used to calculate the likelihood g of a test spectrum as a function of background parameters and source strength. For the test spectrum, g is maximized assuming either the "source presence" hypothesis H_1 or the "background only" hypothesis H_0, and the ratio of the maximized likelihoods under each hypothesis, $R = g_{max}(H_1)/g_{max}(H_0)$, is compared to a predefined threshold set based on an operationally relevant constant false alarm rate.

The PCS algorithm has been applied to other portal and urban detection systems that operate in low signal-to-noise regimes. In [49] and [50], performance enhancement was demonstrated for existing Advanced Spectroscopic Portals (ASP), Standoff Radiation Detection Systems (SORDS) and handheld Radio-Isotope Identifier Devices (RIID) such as Smiths' RadSeeker. However, since the PCS algorithm is proprietary information of Physical Sciences Inc., no details are available in the literature on how the background model is constructed and how the mean background rate as a function of energy as well as the dominant and non-linear modes of spectral variation are calculated. Communication with one of the authors of [46] indicates that these modes are Principal Component Analysis [PCA] describing the background spectra.

5.2.3 Bayesian Aggregation algorithm

In [45] (and more extensively in [44]), another method is presented to account for expected variations in background and common potential nuisance sources causing false alarms in urban searches with mobile spectrometers. Named Bayesian Aggregation (BA), this method relies on field data and injected synthetic sources to learn statistical models of expected threats. BA receives as input radiation spectra and map locations of measurements, and has the three following key stages. The first stage is to estimate the Signal-to-Noise Ratio (SNR) of a measurement in terms of its source signal and background noise components. Once the SNR is estimated, location, velocity, and other positioning information can help quantify the expected exposure to a source. The second stage builds a probabilistic sensor model that can score whether the measurement follows the expected exposure-SNR trend for a point source at hypothetical source locations on the map. Finally, evidence is spatially aggregated across multiple observations using Bayesian data fusion to robustly test these hypothetical source locations and render a threat probability map.

The first phase of the BA algorithm involves training a SNR estimator with single measurements, which in the case of [45] correspond to measurements collected in a field region allocated for "training" in the data. Two methods are discussed for training a measurement SNR estimator: anomaly detectors (based on Principal Components Analysis) and match filters (based on Energy Windowing Regression). The second phase of BA training involves building probabilistic sensor models of expected SNR score distributions as a function of source exposure (i.e., measured source rate). The probabilistic sensor models are estimated from actual (and source-injected) field data, and can be robustly estimated using nonparametric density estimation [51]. The score distribution for negative data (pure background data) forms the null distribution (H_0), and the score distribution obtained for source-injected data becomes the alternate (H_1) probability distribution. The next step of BA is to spatially combine evidence as it is collected. For a given terrain, the scene can be covered with a set of hypothetical source locations; as new measurements are collected and added to the overall data D, BA maintains and updates estimates of the probabilities $P(H_1|D)$ for each source hypothesis and each null hypothesis $P(H_0|D)$.

The authors benchmarked the BA algorithm in comparison to an alternative method of evidence aggregation currently used in the field called the Weighted Combining (WC) Method [52], improving performance. We have not yet reviewed this latter method. We have not found comparison of this method to other methods that do not use data aggregation (i.e., using Bayesian probability aggregation of spectra collected along the mobile detector path). Though this BA method has been prescribed to be used with mobile systems, in the case of a fixed portal inspecting a moving vehicle, spectra before and after the

vehicle occupancy could be aggregated to create a more complete probability density of the null or alternative hypothesis. Note also that this algorithm relies on source injection, and not on source measurements, to generate the alternate (H_1) probability distribution.

5.3 ALGORITHMS ROBUST AGAINST BACKGROUND VARIATIONS FOR PORTAL SYSTEMS

5.3.1 PRINCIPAL COMPONENTS ANALYSIS

The Principal Component Analysis (PCA) of spectral information has already been mentioned as an element of two of the algorithms for mobile systems described above. Here, we present its use in [53] for the spectroscopic analysis of RPM data. In general, this technique is based on anomaly detection, i.e., identifying radiation signatures that are inconsistent with benign sources, which is in contrast to many isotope identification techniques that attempt to identify signatures consistent with specific nuclear threats. In concept, PCA uses eigen analysis of the covariance matrix empirically estimated from the data to make a transformation matrix; multiplication by the transformation matrix transforms the data from a set of interrelated original variables to uncorrelated principal components (PCs) that retain the total variance of a data set, but with a zero-mean shift.

The study in [53] attributes the systematic spectral variability to benign sources given by cargo of ordinary commodities like fertilizer, granite, bananas, etc., or simply to the presence of the intervening cargo. Though not mentioned in [53], weather variation might also contribute to the variability of their data set depending on the time taken to collect the 2000 vehicle spectra used in their analysis. Their PCA result are promising in discriminating anomaly sources from common NORM sources, but are not compared with the performance of other detection algorithms on the same data.

5.3.2 ERNIE

The Enhanced Radiological Nuclear Inspection and Evaluation (ERNIE) tool, presented in [54], is a computer application that uses machine learning analysis to distinguish threats from non-threats in RPM data. Their machine learning model trains on archival RPM scans, extracting available information from the radiation and the vehicle presence sensors. The radiation information includes 8 spectral windows, total counts or intensity, and a spatial distribution of the counts along the inspected vehicle. The model “learns” about potential threats and uncommon sources through augmented RPM scans injected with modelled threat signatures. One aspect relevant to this work is that, according to [54], a “typical expected background” is paired with the measured gamma-ray time profile of each inspected vehicle; however, it is not explicitly described in [54] how the expected background time profile is created.

As output, ERNIE provides the most likely classification of radioactive source type and location, as well as recommended action (Release or Investigate). Machine learning classifies each scan using a Random Forest of decision trees. It is not clear from [54] if the full or the background-subtracted spectrum is used to determine the classification. The effectiveness of ERNIE in identifying most NORM cargo and medical sources reduces secondary inspections by close to an order of magnitude without reduction in sensitivity to other radioactive cargo, while it also provides the source type and location within the inspected vehicle. Since the algorithm is trained with local data, the results are port specific.

5.3.3 OTHER METHODS MENTIONED IN THE LITERATURE

The authors of [46] also list other advanced algorithms that are capable of accurate estimation of the background using information from the entire spectrum: the utilization of ratios of counts within spectral windows to suppress background variability [55], application of principal components analysis and maximum likelihood estimation [47], [53], algorithms that seek to match pre-computed spectral templates [56], [57] and attempts to use neural networks [58].

6 CONSIDERATIONS FOR A FIELD RESPONSE PREDICTION METHODOLOGY

The purpose of this work is to estimate the signal response of a radiation detection system at any field site based on its measured response to a radiation source acquired under controlled conditions at test locations, e.g., the NNSS. To narrow the scope of the problem, we start by focusing on estimating the signal response of a system under test (SUT) at another field site *if measured under exactly the same source configuration* used during testing. Without loss of generality, we assume that the SUT has spectroscopic capabilities, and thus the measured signal is an energy spectrum. The spectrum measured at the NNSS, here denoted as M_{NTS}, when measuring a given source configuration *A*, can be decomposed into a background term B_{NTS} and a source term S_A,

$$M_{NTS} = B_{NTS} + S_A.$$

The spectrum M_{NTS} is the SUT response to the incident source and background radiation, and thus depends on the SUT type and its specific parameters. The label *A* indicates the source contribution to the spectrum according to all parameters that distinguish a given source configuration: source type and strength, distance and position relative to the detector and surroundings, motion with respect to the detector, presence of shielding or other cargo materials and measurement time. If the source term S_A could be accurately estimated, the problem of predicting the spectrum $M_F = B_F + S_A$ that would be obtained if the same source configuration were measured at a field site *F* reduces to estimating the background term B_F at site *F*.

The detector's source response S_A can be isolated in two ways. One conceptually simple way is to measure the background B_{NTS} without the source present and subtract it from M_{NTS} to get $\bar{S}_A = M_{NTS} - B_{NTS}$. Another approach is to estimate S_A through simulation. In order to validate and tune the simulation model, the simulated $\tilde{S}_A$ can be statistically injected on the measured background B_{NTS} and the resultant $\widetilde{M}_{NTS} = B_{NTS} + \tilde{S}_A$ is compared to the measured M_{NTS} [11]. In the modeling of S_A, variations due to changes in the detector operational characteristics that would also affect the source spectral contribution (e.g., due to temperature-induced gain drifts) should be considered. Once estimations of the mean background rate B_F for a field site *F* and of the contribution S_A from a source in configuration *A* are available, they can be Poisson sampled to create a dataset of synthetic measurements $\{M_F\}$ from where ROC curves, or any other relevant figure-of-merit, are generated according to the system's detection algorithm.

Estimating the background B_F will always involve acquiring data representative of the field conditions where the SUT will be deployed. For DNDO, the question is what measurements are necessary and sufficient to validate the SUT detection performance at all of the country's POEs in a manner credible to their stakeholders. Even though in this work we focus on generating a methodology for the prediction of the background B_F to be used in testing, our conclusion as presented in the next section is that predicting an average background is not enough to demonstrate expected SUT performance if the associated background variance at a given field site is large. Besides the statistical Poisson noise, changing background rates can introduce systematic variation in the background distribution expected for a given field site. The correct approach should combine background estimation based on field data collection with a detection algorithm robust against variations within a given range. Thus, the initial goal of estimating the background B_F should be expanded to include the design and demonstration of such detection algorithm.

6.1 ESTIMATING THE BACKGROUND TERM B_F AND ITS VARIANCE

As it has been illustrated in section 3, the radiation background across all country's POEs is expected to vary widely according to its geological location and local scene. To denote such spatial variability, we use the index X so that B_X identifies the background originating from the spatial features of the field site. For example, X could simply label each POE by name. Others have shown ([11]) that it is also possible to group the backgrounds B_X due to the local scene into categories that contain a smaller variance within themselves, so that X could then be a categorical variable taking values like "industrial, rural, downtown, bridges, building interior, open field, etc." Figure 6 shows the background counts of naturally occurring backgrounds lines separated into four distributions according to the local scene [11].

Besides varying by location, the background spectrum of any fixed field site is also expected to change due to daily and seasonal weather variations. To denote those *temporal* variations, we add the index T to the background term $B_{X,T}$. Ideally, numerical weather parameters like temperature, pressure, humidity and time after precipitation would correspond to a background expectation. Categorical variables like time of the day (morning, midday, night) and day in the year or season could also be mapped to an expected background, but doing so would basically amount to using weather parameters. Thus, the background of a given field site labeled above with index F should be thought of as being described by the site's spatial features X and by temporal variations T.

The most accurate prediction of the SUT response $B_{X,T}$ would ideally come from measuring it at each planned deployment site. To account for temporal variations T at a given fixed site X, measurement of $B_{X,T}$ should be performed covering the characteristic span of T for that POE, which would imply measuring at different times of the day and seasons. The results of such an arduous and likely not plausible measurement campaign could be used in several ways. If the spectral variance resulting from T variations is small, a mean spectrum $\langle B_{X,T} \rangle_T$ averaged over the T space could be reported. A measure of the variance of the dataset can be calculated using the total counts in selected energy windows, which for example could correspond with representative background lines (e.g., the main lines from the three primordial decay chains) or with representative SNM lines. The variance of the dataset would be considered small if its square root is comparable to the Poisson noise for the mean $\sqrt{\langle B_{X,T} \rangle_T}$.

If $B_{X,T}$ varies widely over the T space, a library of backgrounds $\langle B_X \rangle_i(T_i)$ could conceivably be created by binning the T space and averaging the measurements over each bin T_i. If the measured spectra $B_{X,T}$ define separate distributions when sorted by categories T_a like, e.g., the time of the day or the day of the year, the library elements could be given by the distributions means $\langle B_X \rangle_a(T_a)$. Cluster analysis techniques [51] applied to the collected background spectra based on some degree of similarity (to be defined) could uncover their dependence on T_i or T_a. We have not found examples in the literature where mapping temporal features with an expected background has been attempted; whether this approach is at all possible remains to be studied with actual data.

It is clear that any one background term B_F to be inserted in $M_F = B_F + S_A$—in order to predict the response at site F—will not cover the full response variability expected in site F, even if measured at the given site. Besides, measuring the background at all deployment sites with the SUT is likely not a practical solution, and mean background spectra according to local scene features will most likely be used instead. Thus, we conclude that, in order to be confident on the predicted performance, the SUT detection algorithm should be robust against the background variance expected at the given field site F.

Although in this work we are limiting our focus to improving the SUT's performance testing phase and not its operation, the two phases are obviously linked. The results of the SUT performance test cannot be separated from its detection algorithm, and therefore, any recommendation for a background-robust algorithm would have to be implemented for the SUT's regular field operation in order to realize the predicted field performance. In section 5.2, several algorithms accounting for the background variability in urban searches with mobile systems were discussed. This manuscript is intended only to frame the problem, and thus, we postpone for the project's next phase the task of developing the appropriate algorithm applicable to portal and handheld systems. What is clear based on the discussion of section 5.2 is that any background-robust algorithm would need to be trained, at a minimum, on data representative of the expected variations at the deployment site.

6.2 UNFOLDING THE BACKGROUND FLUX

In the previous section, the term B_F is assumed to represent the detector's measured response to a radiation background flux incident on the detector from all directions. Hence, B_F not only depends on the incident radiation, but also on the detector features like size, efficiency, resolution, orientation, etc. Hence, background data collected with a detector other than the one undergoing testing, even if both are of the same technology, could contain systematic variations not representative of the field background but of the difference between the two detectors[5].

A solution could in principle be to unfold the incident background flux b_F from the measured response B_F, so that b_F is assumed to be independent of the detection system. A collection of background fluxes $\{b_{X,T}\}$ would then be created to capture the local background variance as described in the previous section. However, the unfolding algorithm might introduce systematic errors beyond those due to the site's

[5] Some algorithms, like GADRAS, can account for detector response differences

background variability. In this approach, the SUT's detector response function would be used to generate an expected background response $\hat{B}_F$ for the SUT at field site F. This constitutes another disadvantage since the SUT response function would either have to be provided by the vendor or modeled as part of the performance testing phase. In either case, experimental data should be used to quantize systematic errors introduced by SUT response function, which could be part of the testing phase at NNSS by using known source and background fluxes. On the other hand, creating a background flux dataset $\{b_{X,T}\}$ that is independent on the detection system allows testing a variety of detector technologies for expected performance at field site F.

6.3 IN SITU BACKGROUND TRAINING AND PREDICTION

An obvious extension of the ideas presented in the previous section is to train the detection algorithm with data collected at the field site where the SUT is deployed. In this regard, the focus of the testing phase could shift from demonstrating how the SUT will perform at particular field sites to validating the ability of the SUT detection algorithm to predict the correct background expectation given sufficient local data.

This approach has numerous advantages. Radiation portal monitors can collect background data continuously with minimal effort. Since these are fixed systems, only the temporal effects, like the daily and seasonal weather variations discussed above or the presence of cargo, would contribute to the background variability. The temporal ambient features would be recorded with extra sensors, like a weather station at each RPM. More importantly, each portal monitor at the same POE should be able to generate its own background expectation according to its surroundings—presence of a nearby wall, exposed to the sun or shade, etc.—, which could be different from the expectation for an identical system positioned differently in the same POE. Moreover, the SUT algorithm could retrain itself frequently with new data, which becomes relevant if the surroundings change.

The in situ data training could be used to generate a prediction of the expected background contribution during a vehicle inspection based on current sensors data. Creating an expectation of the background based on a large sample size dataset that is correlated with the conditions during the actual inspection should increase confidence in source detection, compared to employing a background expectation based on one or just few recent background measurements. Whether it is possible to map the sensors readings T_i to an expected average background spectra $\langle B_X \rangle_i (T_i)$ with small variance still has to be demonstrated, but it constitutes an interesting unsupervised machine learning problem. Principal Components (PC) computed according to local background data, , that estimate the expected background contribution —but not necessarily the full background spectrum—could be used to identify anomalies in the vehicle's measured spectra M_F indicative of the presence a non-background source. The Bayesian Aggregation and the Poisson Clutter Split methods presented above for mobile systems are examples of algorithms that use PC analysis trained with measured data and claim robustness against background variability.

The ERNIE tool presented in section 5.3.2 is one example of supervised machine learning applied to identify nuclear threats. Another example is presented in [59], where classification methods were used and compared in their ability to recognize SNM threats. In these examples, the training set included SNM sources in order to create the various output classes of interest. However, using actual SNM sources is not a practical alternative in field deployments. As done for ERNIE, SNM source terms could be injected into

the locally measured backgrounds in order to create a training set representative of a given SUT environment. Thus, it might then be the job of the DNDO testing phase to create or test the SUT-appropriate synthetic source terms to be used with the supervised learning algorithms.

7 CONCLUSIONS AND RELEVANCE TO DNDO

There is plenty of evidence in the literature that the response of radiation detection systems depends on the conditions at the deployment sites. These external conditions can be separated into fixed site background features, like geological and construction features, and time-varying features, like meteorological conditions. As discussed in this manuscript, variations in external conditions across field sites can produce large variations in the background gamma radiation incident on the detection system, which then produce corresponding variations in the system's spectral response across field sites. Temperature variations during field operation can change the detection system's operational parameters, and thus, change its spectral response.

When testing the performance of a detection system, DNDO testing procedures include exposing the SUT to source materials and configurations of interest. As such, testing will be usually done in locations where relevant quantities of SNM can be used, like the NNSS. The resulting SUT performance is given in terms of the statistical confidence in source detection and rejection, which depends on the estimation of the background contribution during the measurement. However, the background conditions at the testing location will usually not cover the variability in background across the field sites where the detection system will be deployed. This implies that decisions be made based on data from unrealistic environments that are irrelevant for field deployment. Our aim is to assist DNDO in creating testing procedures that make the SUT testing results meaningful for deployment sites.

From the beginning of this work, our intent has been to formulate such testing procedure in the most general way, without prescribing to any particular performance metric used during DNDO testing. Since background estimation is so fundamental to the problem of source detection, we have then focused on the problem of formulating testing procedures that will allow DNDO to generate spectral response data that is representative of the deployment sites. In studying this problem, we have so far reached the following interrelated and preliminary conclusions:

- Estimating the site background will always involve acquiring data representative of the field conditions where the SUT will be deployed.
- Predicting an average background is not enough to demonstrate expected SUT performance if the associated background variance at a given field site is large.
- The correct approach should combine the estimation of the background distribution based on field data collection with a detection algorithm robust against variations within a given range.
- A background-robust algorithm would need to be trained, at a minimum, on data representative of the expected variations at the deployment site.

- A background-robust algorithm used in performance testing would have to be implemented for the SUT's regular field operation in order to realize the predicted field performance.
- An obvious extension of the above ideas is to train the detection algorithm with data collected at the field site where the SUT is deployed. In this regard, the focus of the testing phase could shift from predicting how the SUT will perform at particular field sites to validating the ability of the SUT detection algorithm to predict the correct background expectation given sufficient local data.

We plan to focus the next stage of this work on directly demonstrating the above conclusions with actual data collected with a detector model relevant to DNDO testing and field operations. We will also work towards demonstrating a testing procedure for that specific detector model that produce performance results meaningful for general deployment sites.

8 BIBLIOGRAPHY

[1] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters,* vol. 27, no. 8, pp. 861-874, 2006.

[2] R. C. Runkle, L. E. Smith and A. J. Peurrung, "The photon haystack and emerging radiation detection technology," *Journal of Applied Physics,* vol. 106, no. 4, 2009.

[3] D. J. Mitchell and C. Brusseau, "Neutron Counting and Gamma Spectroscopy with PVT detector," SAND2011-4361.

[4] B. R. S. Minty, "Fundamentals of airbone gamma-ray spectrometry," *AGSO Journal of Australian Geology and Geophysics,* vol. 17, no. 2, pp. 39-50, 1997.

[5] K. E. Marsac et al., "Modeling background radiation using geochemical data: A case study in and around Cameron, Arizona," vol. 165, pp. 68-85, 2016.

[6] B. L. Dickson and K. M. Scott, "Interpretation of aerial gamma-ray surveys—adding the geochemical factors," *AGSO Journal of Australian Geology and Geophysics,* vol. 17, no. 2, pp. 187-200, 1997.

[7] R. B. Firestone and B. Pritychenko, "Nuclear Data Sheets," Tech. Rep. National Nuclear Data Center, 2013.

[8] S. Y. F. Chu, L. P. Ekstrom and R. B. Firestone, "WWW Table of Radioactive Isotopes," 1999. [Online]. Available: http://nucleardata.nuclear.lu.se/toi/.

[9] "Evaluation of Guidelines for Exposure to Technology Enhanced Naturally Occurring Readioactive Materials," [Online]. Available: http://www.nap.edu/catalog/6360.html. [Accessed 19 12 2016].

[10] "United States Geological Survey," [Online]. Available: https://www.usgs.gov/.

[11] T. J. Aucott, *Gamma-Ray background variability in mobile detectors,* PhD. Thesis, U.C. Berkeley, 2004.

[12] R. Pavlovsky et al., "RadWatch Realtime Air monitoring," [Online]. Available: radwatch.berkeley.edu/airsampling.

[13] G. Haquin, "Natural Radioactivity and Radon in Building Materials," [Online]. Available: http://www.irpa12.org.ar/KL/III.4.4/Haquin_fp.pdf.

[14] T. J. Aucott et al., "Routine Surveys for Gamma-Ray Background Characterization," *IEEE Transactions on Nuclear Science,* vol. 60, no. 2, pp. 1147-1150, 2013.

[15] A. Ankney et al., "Muon fluence Measurement for Homeland Security Applications," 2010. [Online]. Available: http://www.pnl.gov/main/publications/external/technical_reports/PNNL-19632.pdf.

[16] P. K. F. Grieder, Cosmic Rays at Earth, Elsevier, 2001.

[17] R. T. Kouzes, J. H. Ely, A. Seifert, E. R. Siciliano, D. R. Weier and L. K. Windsor, "Cosmic-ray-induced ship-effect neutron measurements and implications for cargo scanning at borders," *Nuclear Instruments and Methods in Physics Research A,* vol. 587, pp. 89-100, 2008.

[18] B. Mailyan and A. Chilingarian, "Investigation of diurnal variations of cosmic ray fluxes measured with using ASEC and NMDB monitors," *Advances in Space Research,* vol. 45, no. 11, pp. 1380-1387, 2010.

[19] C. M. Tiwari and D. P. Tiwari, "Characteristics of high energy cosmic ray diurnal anisotropy on day-to-day basis," *Cosmic Research,* vol. 46, no. 5, pp. 465-468, 2008.

[20] G. A. Sandness et al., "Accurate Modeling of the Terrestrial Gamma-Ray Background for Homeland Security Applications," in *IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*, 2009.

[21] G. W. Philips, D. J. Nagel and T. Coffey, "A Primer on the Detection of Nuclear and Radiological Weapons," Center for Technology and National Security Policy, National Defense University, 2005.

[22] R. Pravalie, "Nuclear Weapons Tests and Environmental Consequences: A Global Perspective," *Ambio,* vol. 43, no. 6, pp. 729-744, 2014.

[23] L. J. Mitchell et al., "Cross country background measurements with high purity germanium," in *IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, 2011.

[24] G. F. Knoll, Radiation Detection and Measurement, John Wiley &. Sons, 2010.

[25] E. I. Novikova, B. F. Philips and E. A. Wulf, "A gamma-ray background model for Monte Carlo simulations," *Nuclear Instruments and Methods in Physics Research A,* vol. 579, p. 279–283, 2007.

[26] I. Garishvili et al., "MUSE – MULTI-AGENCY URBAN SEARCH EXPERIMENTS," [Online]. Available: http://web.ornl.gov/~5pe/p102.pdf.

[27] D. E. Peplow, D. Archer and I. Garishvili, "MUSE – MULTI-AGENCY URBAN SEARCH EXPERIMENTS," [Online]. Available: http://web.ornl.gov/~5pe/p999slides.pdf.

[28] "CANBERRA Industries," [Online]. Available: http://www.canberra.com/.

[29] "Ortec," [Online]. Available: https://ortec.com/en-us/.

[30] R. M. Keyser and T. R. Twomey, "Performance of a Radiation Portal Freight Monitor based on Integrated Germanium Detector Modules," in *INMM 2010 Annual Meeting*, 2010.

[31] C. Hannigan, *Investigation of temperature effects on electro-mechanically cooled Ge detectors,* Masters Thesis, Department of Physics, University of York, 2010.

[32] P. L. Reeder and D. C. Stromswold, "Performance of large NaI(Tl) gamma-ray detectors over temperature -50 0C to 60 0C," PNNL-14735, 2004.

[33] K. D. Ianakiev et. al., "Temperature behavior of NaI(Tl) scintillation detectors," *Nuclear Instruments and Methods in Physics Research A,* vol. 607, pp. 432-438, 2009.

[34] B. Alexandrov, "Temperature Behavior of the doped NaI(Tl) Scintillators and its impact on the pulse height analysis instrumentation," LA-UR-05-4277, 2005.

[35] J. S. Schweitzer and W. Ziehl, "Temperature dependence of NaI(Tl) decay constant," *IEEE Transactions on Nuclear Science,* vol. 30, no. 1, February 1983.

[36] R. T. Kouzes, E. R. Siciliano, J. H. Ely, P. E. Keller and J. R. J. McConn, "Passive Neutron Detection for Interdiction of Nuclear Material at Borders," *Nuclear Instruments and Methods in Physics Research A,* vol. 584, pp. 383-400, 2008.

[37] K. D. Jarman et al., "A comparison of simple algorithms for gamma-ray spectrometers in radioactive source search applications," vol. 66, pp. 362-371, 2008.

[38] K. P. Ziock et al., "Source-search sensitivity of a large-area, coded aperture, gamma-ray imager," *IEEE Transactions on Nuclear Science,* vol. 53, no. 3, pp. 1614-1621, 2006.

[39] K. P. Ziock and K. E. Nelson, "Maximum detector sizes required for orphan source detection," *Nuclear Instruments and Methods in Physics Research A,* vol. 579, pp. 357-362, 2007.

[40] T. J. Aucott et al., "Impact of detector efficiency and energy resolution on gamma-ray background rejection in mobile spectroscopy and imaging systems," *Nuclear Instruments and Methods in Physics Research A,* vol. 789, pp. 128-133, 2015.

[41] S. Agostinelli et al., "Geant4—a simulation toolkit," *Nuclear Instruments and Methods in Physics Research A,* vol. 506, no. 3, p. 250, 2003.

[42] MCNP X-5 Monte Carlo Team, "MCNP – A General Purpose Monte Carlo N-Particle Transport Code, Version 5," Los Alamos National Laboratory, Los Alamos, NM, 2003.

[43] T. J. Aucott et al., "Effects of Background on Gamma-Ray Detection for Mobile Spectroscopy and Imaging Systems," *IEEE Transactions on Nuclear Science,* vol. 61, no. 2, April 2014.

[44] P. Tandon, *Bayesian aggregation of Evidence for Detection and characterization of patterns in multiple noisy observations,* Ph.D. Thesis, Robotics Institute, Carnegie Mellon University, 2015.

[45] R. Tandom, "Detection of radioactive sources in urban scenes using Bayesian Aggregation of data from mobile spectrometers," *Information systems,* vol. 57, pp. 195-206, 2016.

[46] T. H. Joshi et al., "A Comparison of the detection sensitivity of the Poisson Clutter Split and the Region of Interest Algorithms on the RadMAP Mobile system," *IEEE Transactions on Nuclear Science,* vol. 63, no. 2, April 2016.

[47] S. E. Labov et al., "Foundations for Improvements to Passive Detection Systems - Final Report," October 2004. [Online]. Available: https://e-reports-ext.llnl.gov/pdf/312592.pdf.

[48] T. Burr and M. Hamada, "Radio-isotope identification algorithms for NaI spectra," *Algorithms,* vol. 2, no. 1, p. 339, 2009.

[49] B. R. Cosofret, K. N. Shokhirev, P. A. Mulhall, D. Payne and B. Harris, "Utilization of advanced clutter suppression algorithms for improved standoff detection and identification of radionuclide threats," in *2014 SPIE DSS*, 2014.

[50] B. R. Cosofret, K. N. Shokhirev, P. A. Mulhall, D. Payne, B. Harris, E. Arsenault and R. Moro, "Utilization of advanced clutter suppression algorithms for improved spectroscopic portal capability against radionuclide threats," in *IEEE International Conference of Technologies for Homeland Security*, Waltham, MA, 2013.

[51] T. Hastie, R. Tibshirani and J. Friedman, The Elements of Statistical Learning, Second Edition, Berlin: Springer, 2009.

[52] K. Nelson and S. Labov, "Aggregation of Mobile Radiation Data," Lawrence Livermore National Lab Technical Report.

[53] R. C. Runkle, M. F. Tardiff, K. K. Anderson, D. K. Carlson and L. E. Smith, "Analysis of spectroscopic radiation portal monitor data using principal components analysis," *IEEE Transactions on Nuclear Science,* vol. 53, no. 3, p. 1418–1423, June 2006.

[54] S. Labov, "Enhanced Radiological Nuclear Inspection and Evaluation (ERNIE), OUO," Lawrence Livermore National Lab, 2014.

[55] R. S. Detwiler et al., "Spectra anomaly methods for aerial detection using KUT nuisance rejection," *Nuclear Instruments and Methods in Physics Research A,* vol. 784, pp. 339-345, 2015.

[56] E. LaVigne, G. Sjoden, J. Baciak Jr. and R. Detwiler, "Extraordinary improvement in scintillation detectors via post-processing with asedra: solution to a 50-year-old problem," in *Proceedings of SPIE - The International Society for Optical Engineering*, 2008.

[57] R. J. Estep, C. W. McCluskey and B. A. Sapp, "The multiple isotope material basis set (mimbs) method for isotope identification with low- and medium-resolution gamma-ray detectors," *Journal of Radioanalytical and Nuclear Chemistry,* vol. 276, no. 3, p. 737–741, 2008.

[58] P. Olmos, J. Diaz, J. Perez, G. Garcia-Belmonte, P. Gomez and V. Rodellar, "Application of neural network techniques in gamma spectroscopy," *Nuclear Instruments and Methods in Physics Research A,* vol. 312, no. 1, p. 167–173, 1992.

[59] Y. Yang, *Machine Learning Techniques in Nuclear Detection, Drug Ranking and Video Tracking,* Ph.D. Thesis, Industrial Engineering and Operations Research, UC Berkeley, 2013.